基于笔画角度变换和宽度特征的自然场景文本检测 *

陈　硕 [1,2]，郑建彬 [1,2]，詹恩奇 [1,2†]，汪　阳 [1,2]

(1. 武汉理工大学 信息工程学院，武汉 430070; 2. 光纤传感技术与信息处理教育部重点实验室，武汉 430070)

摘　要：针对光照不均和背景复杂度所导致的自然场景文本检测中文本的漏检和错检现象，提出一种基于笔画角度变换和宽度特征的自然场景文本检测方法。分析发现与非文本相比，文本具有较稳定的笔画角度变换次数和笔画宽度，针对这两个特性提出笔画外边界优劣角变换次数和增强笔画支持像素面积比两种特征。前者分段统计笔画外轮廓角度变换次数；后者计算笔画宽度稳定区域在笔画总面积的占比，用来分别反映笔画角度和宽度变化稳定特性。为降低文本漏检率，采用多通道最大稳定极值区域（maximally stable extremal regions，MSER）检测，合并所有候选区域，提取候选区域的笔画特征和纹理特征，利用支持向量机完成文本和非文本区域分类。在ICDAR2015数据库上，算法的精确率和召回率分别达到79.3%和72.8%，并在一定程度上解决了光照不均和复杂背景的问题。

关键词：自然场景；文本检测；笔画特征

中图分类号：TP391　　doi: 10.3969/j.issn.1001-3695.2017.10.1019

Text detection based on stroke angle conversion and stroke width features in natural scene

Chen Shuo[1,2], Zheng Jianbin[1,2], Zhan Enqi[1,2†], Wang Yang[1,2]

(1. *College of Information Engineering*, *Wuhan University of Technology*, *Wuhan* 430070, *China*; 2. *Key Laboratory of Fiber Optic Sensing Technology and Information Processing of Ministry of Education*, *Wuhan* 430070, *China*)

Abstract: In order to reduce the missing detection and misclassification of text caused by uneven illumination and background complexity in text detection of natural scenes, this paper presented a natural scene text detection method based on stroke angle transformation and width features. Compared to non-text, the text has a more stable performance of stroke outline angle conversion times and stroke width. Therefore, this paper proposed methods of extracting the number of transformations of the outer corner of the stroke and the enhancement of the pixel area ratio of the stroke support. In order to extract the characteristics of angular conversion, it used the method of outer contour segmentation to calculate the number of conversion times. In order to extract the strokes width characteristics, it calculated the proportion of the width stable area in the total strokes area. To reduce rate of the text missing detection, multi-channel MSER was used to detect text candidate area. Candidate areas in all channels were merged to extract the stroke and texture features. Support vector machines combined with features adopted, it used to classify text and non-text. The simulations show that the accuracy and recall rate of the algorithm were 79.3% and 72.8% in the ICDAR2015 database, respectively. Moreover, it solves the problem of uneven illumination and complex background to some extent.

Key Words: natural scene; text detection; stroke feature

0　引言

自然场景图像是指取材于真实生活场景的图像，蕴涵许多重要的文本信息，如商品包装、道路指示牌、宣传标语等，这些文字含有明确语义信息，更好地表达视觉场景。据统计，人们每天获取的信息约有 70%来自视觉观察和感知[1]，故从图片中提取文本信息将有效地提高信息交流的速度，而自然场景文本检测作为提取文本信息提取的基础，具有极大的研究价值。

造成自然场景文本检测难度较大的原因有以下三点：a)不同场景下的文本多是不可控的，它们具有不同的颜色、字体、尺度、方向等；b)自然场景下的背景更为复杂，会出现与文本特征相似的物品（如栏杆、砖块等）对文本检测造成干扰；c)拍摄过程导致的图像模糊、部分文本被遮挡、光照不均等问题也会为文本检测增加难度。

基金项目：国家自然科学基金资助项目（61303028）

作者简介：陈硕（1993-），女，硕士研究生，主要研究方向为模式识别、图像处理；郑建彬（1966-），男，教授，博士，主要研究方向为模式识别、嵌入式系统；詹恩奇（1972-），男（通信作者），副教授，博士，主要研究方向为信号处理、模式识别（eqzhan@whut.edu.cn）；汪阳（1977-），男，副教授，博士，主要研究方向为机器人控制、嵌入式系统.

chinaXiv:201805.00026v1

ChinaXiv合作期刊

按照提取文本候选区域的方式，目前自然场景文本检测的方法主要分为基于滑动窗口（sliding-window）的方法和基于连通域（connected-component）的方法两类[2~4]。基于滑动窗口的方法主要利用的是文本区域与非文本区域有不同的纹理特征，如局部强度、滤波器响应、小波系数等特征[5]。此类方法虽然对于噪声和光照稳定，但无法完整地检测出单个文本，过多的滑动窗口令算法的计算效率较低，故只适用于简单的场景图片。

目前常用的文本检测方法为基于连通域的方法，此方法利用了文本具有相同的颜色和纹理特征等特性，通过一系列连通域分析法提取文本区域。常用的文本区域提取方法有极值区域（ER）[6]算法、最大稳定极值区域（MSER）[7]算法和笔画变换（SWT）[8]算法。这些方法虽然对于光照和噪声较为敏感，但是在检测速度上有极大的提升，通过调整检测参数基本上可以检测出所有文本区域，且大多数文本候选区域为单个文本。Yin等人[9]选取空间距离、宽和高度差、顶部连线倾斜角度、底部连线倾斜角度、颜色差、笔画宽度差组成特征空间，用字符分类器估计非文本的后验概率，消除概率高的非文本区，最后使用贝叶斯分类器对文本分类。该算法获得较高的准确率，但是对于背景复杂的图片则会遗漏大量的文本区。

针对自然场景图像中光照和复杂背景等对文本检测影响较大的因素，本文提出一种基于笔画特征和多种特征融合的自然场景文本检测方法。该算法的流程如图 1 所示。首先将自然场景图像进行多通道候选文本区域提取并去重；然后取候选文本区域的二值图像提取笔画特征，通过启发式规则滤除一部分非文本区；接着采用梯度方向直方图（histogram of oriented gradient，HOG）特征、平均局部二值模式（mean local binary pattern，MLBP）特征结合支持向量机（support vector machine，SVM）对剩下的候选区域进行分类；最后利用利用广度优先搜索的思想实现文本行聚合得到文本检测结果。

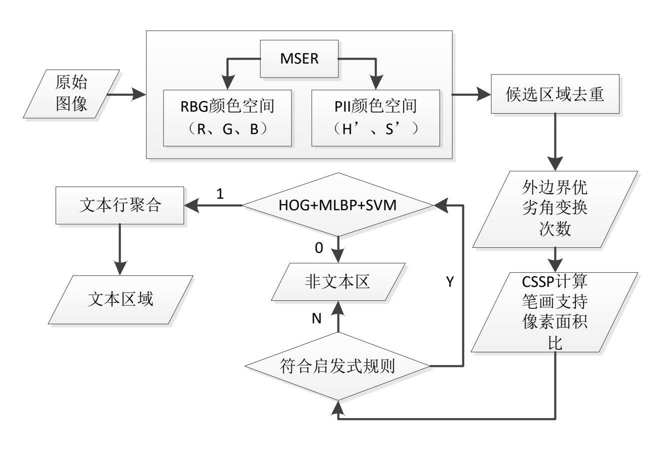


图 1　文本算法流程

1　多通道 MSER 的文本候选区域提取和去重

鉴于传统的 MSER 算法对于光照和噪声较为敏感，故对原始自然图像使用双边滤波进行预处理消除噪声影响，再提取多个颜色通道的 MSER 作为文本候选区域来减小光照不均的干扰。将滤波后的 RGB 图像转换至基于感知的光照不变（perception-based illumination invariant, PII）颜色空间[10]。 PII 颜色空间的优势在于：a）不同光照条件下，任意两种颜色的差值向量基本保持不变；b）任意两种颜色的欧氏距离和人眼视觉感知距离基本一致[11]。为检验不同通道组合的效果，从 ICDAR2015 数据集中随机抽取 50 张图片进行测试，以有效区域比例和文本覆盖率两项指标进行对比，所得结果如表 1 所示。

表 1　不同通道组合 MSER 检测结果

通道	有效区域比例/%	文本覆盖率/%
gray	64.42	85.33
R+G+B	65.23	89.65
gray+R+G+B	64.94	90.14
H'+S'	65.78	83.45
R+G+B+H'+S'	65.57	93.29

由表 1 可知，选择 R、G、B 通道和 H'、S'通道结合可获得最大的文本覆盖率，即可以最大程度地检测出图像中的文本区域，且有效区域的比例较高。

多层 MSER 算法检测会导致颜色稳定区域被多次检测，会产生大量重复的文本候选区域。为了减少无效计算，进行去重复处理。此处选择通过面积重叠率 $\psi(R_{i,j})$ 和区域颜色变化率 var 判断两个文本候选区域是否重合，计算公式如下：

$$\psi(R_{i,j}) = \frac{A(R_i) \cap A(R_j)}{A(R_i) \cup A(R_j)} \tag{1}$$

$$\text{var} = norm(std(R, G, B)) \tag{2}$$

式（1）中，$A(R_x)$ 表示某文本候选区域 R_x 中 MSER 的像素个数；R, G, B 分别表示某区域在 RGB 颜色空间的三个分量通道的值；式（2）中，$std(x)$ 表示求向量 x 的标准差，$norm(y)$ 表示求向量 y 的 L2 范数。若面积重叠率 $\psi(R_{i,j}) \approx 1$，则去除其他区域，只保留其中颜色变换率 var 最小的一个文本候选区。去重复效果如图 2 所示。

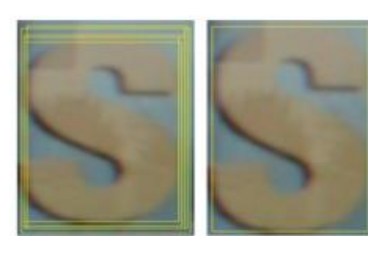


图 2　去重复效果

2　基于笔画特征的文本检测

提取的文本候选区域不仅尺寸差异较大，还具有不同的宽高比 $A_s = w / h$，其中 w, h 为文本候选区域外接矩形的宽和高。

chinaXiv:201805.00026v1

ChinaXiv合作期刊

笔画特征是基于候选区域均为单字符提取的，据统计，单个字符一般 $A_s \in [0.3,1.5]$，宽高比较大或较小的区域大多数是非文本区，也可能是多个字符连接在一起。对于宽高比不属于 $[0.3,1.5]$ 内的候选区域，计算笔画特征时需要引入权重 $N=[\max\{A_s,\frac{1}{A_s}\}]$，其中 $[\bullet]$ 表示四舍五入取整，物理意义为候选区域可分割成宽高比为 1 的区域的个数。

2.1 笔画特征计算

提取的文本候选区域中不仅包括文本区，还含有大量的非文本区，如何准确地滤除非文本区是本节主要解决的问题。与非文本相比，文本具有明显的笔画特征，如清晰的笔画轮廓和稳定的笔画宽度等，文本选取以下两个特征——外边界优劣角变换次数和增强笔画支持像素面积比，作为笔画特征用于分类。

2.1.1 外边界优劣角变换次数

26 个英文字母均由有限数目笔画构成，有限稳定笔画构成有限外边界角点，而外边界角点分为优角和劣角，如图 3 所示。



图 3　二值图像外边缘角点分布和轮廓图

图中红色的点为优角，绿色的点为劣角（见电子版），外边缘上优角和劣角的变换次数为 6。

由于二值图像非 1 即 0，其边缘梯度较大，且轮廓不够平滑，使用 Harris 角点检测会误检出多余角点且对于接近于平角的角点不够敏感，所以针对于二值图像，提出一种新的角点检测方法。

先提取二值图像的边缘像素，将边缘像素按照四领域像素分布分为 8 类，依次编号为 1 到 8，分类如图 4 所示。

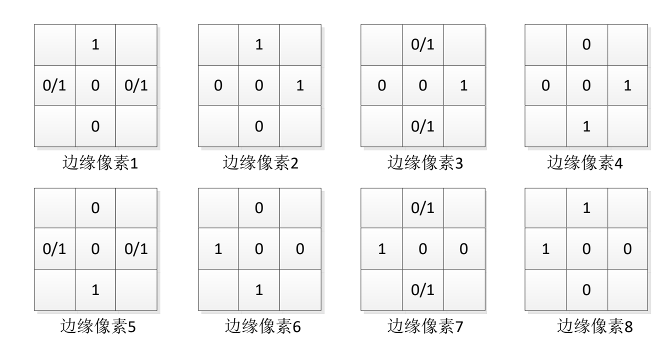


图 4　边缘像素分类图

图 4 中 0/1 表示一个边缘像素类型中这两个位置的像素必须同为 0 或同为 1。以外轮廓右上角的边缘像素为起点，沿顺时针方向标记每个边缘像素的类型，标记值构成一维向量。接着对一维向量进行分段，令每个分段点对应外边缘上的角点。

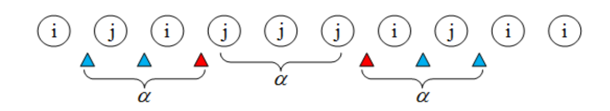
分段步骤如下：

a）两个跳变点之间的像素个数超过阈值 α，将两端标记值跳变点转换为候选分段点，如图 5（a）所示。

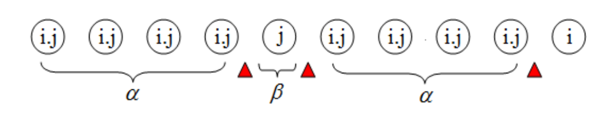
b）连续 m 个（$m > \alpha$）标记值跳变点（包括候选分段点）前后的标记值均为标记值 i 和标记值 j，则这若干个标记值跳变点之间的标记值合并为一个段向量，记为段向量 $i.j$，其间的标记值也转换为 $i.j$（因为在此情况下边缘像素 1、3、5、7 不可能相邻，边缘像素 2、4、6、8 不可能相邻，故 i 和 j 必为一个奇数一个偶数，在此令 i 为奇数，j 为偶数），如图 5（b）所示。

c）重复过程 a)和 b)，直到不能合并。

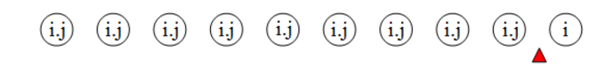
d）两段 $i.x$（x 取值 0-8）之间的间隔像素小于等于阈值 β，则这两个段向量合并为一个段向量 $i.x$，其他的跳变点直接转换为候选分段点，如图 5（c）所示。



(a)原始数据排列



(b) i.j 组合后数据排列



(c)最终分段结果

图 5　边缘分段过程

图 5 中，蓝色三角为标记值跳变点，红色三角为候选分段点（见电子版），此时标记值构成的一维整数向量变成了由段向量构成的一维小数向量，段向量的数值和段向量代表的轮廓方向如图 6 所示。

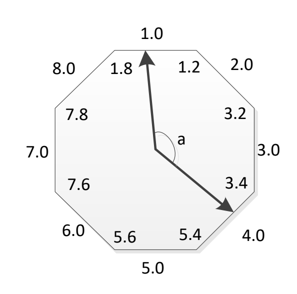


图 6　段向量数值与对应轮廓方向

若分段点前后的数值按顺时针方向变化（保证夹角 a 小于 180°），则此分段点对应优角；若按逆时针方向变化，则对应劣角，其变换次数记为 k。记当前的分段点为 P，分段点之前的标记值为 $i_p.j_p$，分段点之后的标记值为 $i_{p+1}.j_{p+1}$，计算公式如下：

chinaXiv:201805.00026v1

$$\Delta_i = i_{p+1} - i_p, \Delta_j = j_{p+1} - j_p$$

$$P = \begin{cases} 1, \Delta_i \in [1,3] \cup [-5,-7] 或 \Delta_i \in \{0,4\} \cap \Delta_j \in [1,3] \cup [-5,-7] \\ -1, \Delta_i \in [-1,-3] \cup [5,7] 或 \Delta_i \in \{0,4\} \cap \Delta_j \in [-1,-3] \cup [5,7] \\ \varnothing, \quad \Delta_i \in \{0,4\} \cap \Delta_j \in \{0,4\} \end{cases} \tag{3}$$

式（3）中 $\Delta_i, \Delta_j \in \{0,4\}$ 这种情况是不存在的；k 为 P 正负值跳变的次数。对于 $A_s \notin [0.3,1.5]$ 的候选区域，$k = k / N$。阈值 α 和 β 对 k 的影响如表 2 所示。

表 2　不同阈值对 k 的影响

α	β	文本样本 k 的最大值	>k 的非文本样本比例/%
	1	19	65.7
2	2	15	66.2
	3	-	-
	1	14	67.8
3	2	10	68.3
	3	9	63.7
	1	23	64.5
4	2	16	65.6
	3	11	64.1

由表 2 可知，取 $\alpha = 3, \beta = 2$ 时，去除 $k > 10$ 的文本候选区域效果最好。

2.1.2 增强笔画支持像素面积比（ESSP）

使用笔画支持像素（stroke support pixel, SSP）[7]法计算笔画骨架，定义笔画骨架距离边缘像素的距离为 $d_i = S_w / 2$，S_w 为笔画宽度，笔画骨架的长度为 L，对于理想的笔画，笔画宽度基本保持稳定，如图 7 所示，笔画支持像素的个数 $A_s = S_w \times L = 2\sum_{i \in S} d_i$ 基本上等于笔画面积。

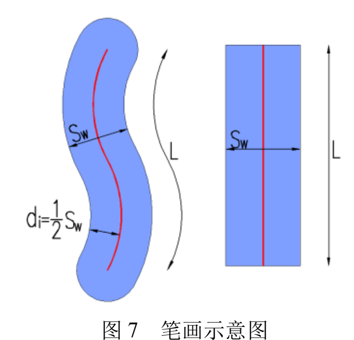


图 7　笔画示意图

计算笔画骨架需要先计算距离变换图，即二值图像上每一个像素点的值是它离其最近背景点的距离，选取当前像素的距离值大于等于其八邻域的像素点作为笔画骨架像素。一般状况下，会出现笔画骨架不连续、一行有多个笔画骨架像素的等情况，简单计算会造成结果不准确。因此，为每个骨架像素引入一个权重系数 w_i，将其归一化到单一的笔画长度，新的支持像素个数计算公式为

$$A_s = 2\sum_{i \in S} w_i d_i \tag{4}$$

$$w_i = 3 / N_i \tag{5}$$

式（5）中 N_i 为 3×3 邻域内骨架像素的个数，笔画面积比 $\xi = A_s / A$，其中 A 为二值区域面积，计算过程如图 8 所示。

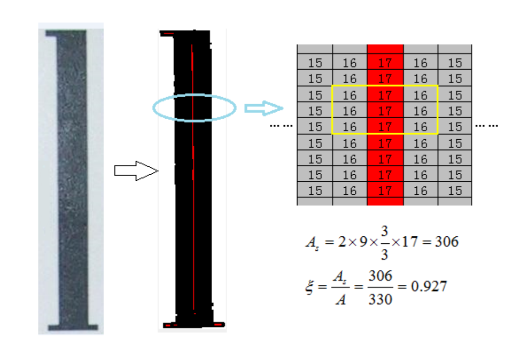


图 8　笔画面积比计算过程

即使加入权重系数，对于笔画宽度发生变化的文本，ξ 值较小，产生这一影响的根本原因是缺失了太多的骨架像素。当笔画宽度由小到大变化或多个笔画相连时，骨架像素的距离值不再是局部最大值，因此无法被检测出，如图 9（a）所示。

为解决这一问题，提出增强笔画支持像素（enhance stroke support pixel, ESSP）的方法。对于 $A_s \notin [0.3,1.5]$ 的候选区域，首先对候选区域的较大边缘进行缩放，保证候选区域能被平均分割，缩放率 $ratio = N / \max\{A_s, 1/A_s\}$，再将候选区域沿较大边缘分割成 N 份子候选区域，接着对子候选区域进行上述计算，得到各个子候选区域的骨架像素，如图 9（b）所示。

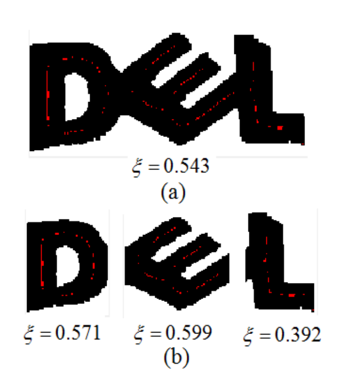


图 9　SSP 检测骨架像素（a）分割后检测骨架像素（b）

ESSP 方法的计算过程如下：

a）计算骨架像素距离的平均值 $d = \frac{1}{n}\sum_{i=1}^{n} d_i$。

b）选择笔画骨架上像素距离接近于 d 的端点像素（包括 3×3 邻域内只有一个骨架像素的点）作为原点，寻找据原点最近的两个骨架像素，用直线分别连接原点和两个像素。

c）若直线上存在背景像素，则放弃此条直线，跳转至下一个原点，重复步骤 b)；若不存在背景像素，且两个像素点的距离值之差小于 $0.3d$，则选取直线上像素 3×3 邻域的最大值（若最大值为骨架或位于直线上则选次大值）加入骨架像素，否则

chinaXiv:201805.00026v1

chinaXiv:201805.00026v1

放弃此条直线，重复步骤 b)。

d）找回所有骨架像素后，使用公式 $A_s = 2\sum_{i \in S} w_i d_i$ 和 $\xi = A_s / A$ 计算像素面积比。结果有较大的改善，计算过程如图 10 所示。

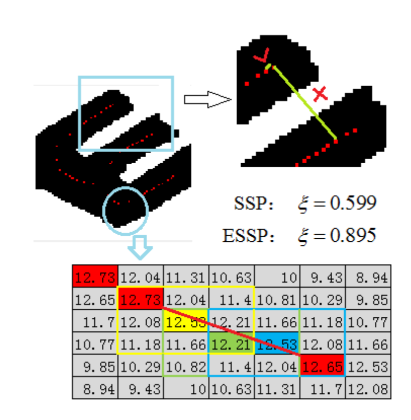


图 10　ESSP 计算过程

任意选取各取 500 张文本样本图和非文本样本图进行算法检测，SSP、ESSP 的对比结果如图 11 所示。

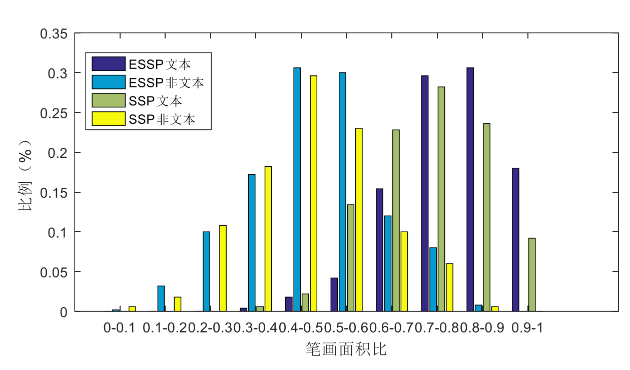


图 11　SSP 和 ESSP 的对比结果

图 11 中 0.1-0.2 表示笔画面积比位于区间 $(0.1, 0.2]$，其他数值以此类推。由图可见，对于文本样本，ESSP 有效地增大的笔画面积比；对于非文本样本也有所增强，但增强的幅度不大。使用 ESSP 算法增大了文本和非文本的差距。

2.2　文本检测

文本和非文本不仅在笔画特征上有差别，同时文本具有更为规律的轮廓和稳定的边缘及纹理特征，本文选用 HOG 特征、MLBP 特征和 SVM，滤除剩下的非文本区。HOG 通过计算和统计图像局部区域的梯度方向直方图来构成特征向量，描述图像的梯度方向和边缘信息。MLBP[12]是对局部二值模式（local binary pattern，LBP）的一种改进，提升了算法对噪点的鲁棒性。

上述步骤去除了大多数复杂的背景元素，得到了相对完整的单个字符区域。接下来利用广度优先搜索(breadth first search，BFS)的思想，对这些单个字符区域进行聚合形成文本行，并以此作为本文自然场景文本定位的最终结果。BFS 是连通图的一种遍历方法，从一个顶点 v_0 开始依次访问其邻接点 $v_1, v_2 \cdots v_k$，若特征相似，则合并成一个一个顶点，接着访问 $v_1, v_2 \cdots v_k$ 的邻接点，直到遍历图中所有顶点。本文选取的特征有 CIE94 颜色距离、字符空隙距离，中心连线倾斜角度[13]、笔画宽度作为文本区域聚合的限制条件。将颜色距离和笔画宽度相近的文本区域合并为一个组，一一统计组中各文本区域之间的字符空隙距离和中心连线倾斜角度，并根据相似度合并文本区域形成文本行。

3　实验结果分析

3.1　实验数据库和评价指标

选择公开数据库 ICDAR2015 数据集对算法进行检测。ICDAR2015 数据库包含不同颜色、字体、尺寸的文本图像，其中部分图像具有光照不均、对比度低、模糊、遮挡等增问题，其中训练集 229 张图片，测试集 233 张图片。

本文采取 ICDAR 提供评价指标对本文提出方法进行评估，评价指标由精确率 P（precision Rate）、召回率 R（recall Rate）和综合评价值 F 值组成[14]。其定义如下：

$$m(r;R) = \max\{m_p(r;r') \mid r' \in R\} \tag{6}$$

$$P = \frac{\sum_{r_e \in E} m(r_e, T)}{|E|} \tag{7}$$

$$R = \frac{\sum_{r_t \in T} m(r_t, T)}{|T|} \tag{8}$$

其中：图片中真实文本区域的外接矩形集合为 T；算法输出的文本区域外接矩形集合为 E；$m_p \in [0,1]$ 为面积重合率，计算方式如式（1）所示，不同的是 $A(R_x)$ 表示文本区外接矩形的面积。

将精确率 P 和召回率 R 结合得到综合评价指标 F，一般取 $\alpha = 0.5$，计算公式如下：

$$F = \frac{1}{\alpha / P + (1-\alpha) / R} \tag{9}$$

3.2　结果分析

ICDAR 公开数据库是文本检测常用数据集，部分结果展示如图 12 所示。



图 12　部分结果展示

从图 12 可以看出，在本文方法能够适应不同背景、不同字

ChinaXiv合作期刊

体大小、不同文本颜色等情况，有效地抑制背景中类似文本的噪声，如树叶、箭头、窗户等干扰，准确定位出文本位置，能得到较好的检测结果。

为进一步验证算法的有效性，将算法的精确率 P、召回率 R 和综合评价指标 F 与国内外先进算法进行对比，对比结果如表 3 所示。

表 3 相关算法性能比较

算法	R/%	P/%	F/%
Shi's Method [15]	62.9	84.7	72.2
Bais's Method [16]	68.2	78.9	73.2
Neumann's Method [6]	64.8	87.5	74.5
Yi's Method [9]	66.5	88.5	75.9
本文算法	72.8	79.3	75.9

从表 3 可知，本文算法在 ICDAR2015 数据集上测试，虽然精确度 P 较低，仅超过 Bais 的算法，但是 F 值与 Yi 的方法相同，高于其他算法，且具有最高的召回率 R。为了尽可能地保存更为完全的真正文本区域，即提高召回率，本文算法提取候选区域时采用多通道 MSER 算法，且所提取的笔画特征也是针对文本和非文本的本质区别，剔除了大量非文本区，最大限度地筛选出真实文本，但同时也保留了具有笔画特征的人造图形，进一步使用边缘特征和纹理特征区分文本和非文本，由此导致部分非文本被误判为文本，使精确度偏低。综合几个评估标准，本文算法具有较好的性能。

4 结束语

本文提出一种基于笔画特征的自然场景文本检测方法。对多通道 MSER 算法进行去重处理，既克服了传统 MSER 算法对光照的敏感性，又抑制了候选区域数量的暴增。提取候选区域的笔画特征即包含了笔画的轮廓特征和骨架特征，可以从本质描述文本和非文本的区别，综合边缘特征和纹理特征，能较好地检测自然场景中的文本区域。但是对于对比度较低、模糊度较高、文本被栏杆遮挡的自然场景图片，本文算法不能准确定位，还需进一步研究，提升算法性能。

参考文献：

[1] 许晓丽. 基于聚类分析的图像分割算法研究 [D]. 哈尔滨: 哈尔滨工程大学, 2012.

[2] Jaderberg M, Simonyan K, Vedaldi A, et al. Reading text in the wild with convolutional neural networks [J]. International Journal of Computer Vision, 2016, 116 (1): 1-20.

[3] Huang Weilin, Qiao Yu, Tang Xiaoou. Robust scene text detection with convolution neural network induced MSER trees [C]// Proc of European Conference on Computer Vision. 2014: 497-511.

[4] Zhang Zheng, Shen Wei, Yao Cong, et al. Symmetry-based text line detection in natural scenes [C]// Proc of IEEE Conference on Computer Vision & Pattern Recognition. 2015: 2558-2567.

[5] Zhu Yingying, Yao Cong, Bai Xiang. Scene text detection and recognition: recent advances and future trends [J]. Frontiers of Computer Science, 2016, 10 (1): 19-36.

[6] Neumann L, Matas J. Real-time scene text localization and recognition [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2012: 3538-3545.

[7] Neumann L, Matas J. Efficient scene text localization and recognition with local character refinement [C]// Proc of International Conference on Document Analysis and Recognition. 2015: 746-750.

[8] Xu Hailiang, Xue Like, Su Feng. Scene text detection based on robust stroke width transform and deep belief network [C]// Proc of Asian Conference on Computer Vision. 2014: 195-209.

[9] Yin Xucheng, Yin Xuwang, Huang Kaizhu, et al. Robust text detection in natural scene images [J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2014, 36 (5): 970-983.

[10] Chong H Y, Gortler S J, Zickler T. A perception-based color space for illumination-invariant image processing [J]. ACM Trans on Graphics, 2008, 27 (3): 1-7.

[11] 孙雷. 自然场景图像中的文字检测 [D]. 合肥: 中国科学技术大学, 2015.

[12] Cho H, Sung M, Jun B. Canny text detector: fast and robust scene text localization algorithm [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2016: 3566-3573.

[13] Buta M, Neumann L, Matas J. FASText: efficient unconstrained scene text detector [C]// Proc of IEEE International Conference on Computer Vision. 2015: 1206-1214.

[14] LUCAS S M. ICDAR 2005 text locating competition results [C]// Proc of International Conference on Document Analysis and Recognition. 2005: 80-84.

[15] Shi Cunzhao, Wang Chunheng, Xiao Baihua, et al. Scene text recognition using part-based tree-structured character detection [C]// Proc of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2013: 2961-2968.

[16] Bai Bo, Yin Fei, Liu Chenglin. Scene text localization using gradient local correlation [C]// Proc of International Conference on Document Analysis and Recognition. 2013: 1380-1384.

chinaXiv:201805.00026v1